AMPLITUDE AND PHASE MATCHING FOR LAYERED MODULATION RECEPTION

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This is a continuation-in-part application and claims the benefit under 35 U.S.C. Section 120 of the following co-pending and commonly-assigned U.S. utility patent application, which is incorporated by reference herein:

[0002] Utility Application Serial No. 09/844,401, filed April 27, 2001, by Ernest C. Chen, entitled "LAYERED MODULATION FOR DIGITAL SIGNALS," attorneys' docket number PD-200181 (109.51-US-01).

[0003] This application claims the benefit under 35 U.S.C. §119(e) of the following U.S. Provisional Patent Application, which is incorporated by reference herein:

[0004] Application Serial No. 60/421,332, filed October 25, 2002, by Ernest C. Chen, Jeng-Hong Chen, Kenneth Shum and Joungheon Oh, entitled "AMPLITUDE AND PHASE MATCHING FOR LAYERED MODULATION RECEPTION," attorneys' docket number PD-201033 (109.73-US-P1).

BACKGROUND OF THE INVENTION

1. Field of the Invention

[0005] The present invention relates to systems and methods for receiving layered modulation signals, particularly in a direct satellite broadcast system.

2. Description of the Related Art

[0006] Digital signal communication systems have been used in various fields, including digital TV signal transmission, either terrestrial or satellite. As the various

digital signal communication systems and services evolve, there is a burgeoning demand for increased data throughput and added services.

[0007] It has been proposed that a layered modulation signal, transmitting coherently or non-coherently both upper and lower layer signals, can be employed to meet these needs and more. Such layered modulation systems allow higher information throughput, with and without backward compatibility. When backward compatibility is not required (such as with an entirely new system), layered modulation can still be advantageous because it requires a TWTA peak power significantly lower than that for a conventional 8PSK or 16QAM modulation format for a given throughput.

[0008] However, to receive such layered modulation signals requires reconstruction of the upper layer signals to remove them from the total signal for lower layer signal processing to occur. Further, the performance of lower layer demodulation depends on the cancellation accuracy. The reconstructed signal should optimally match the received signal in overall amplitude and phase. Therefore, amplitude and phase errors in the reconstructed signal at the point of signal cancellation need to be estimated.

[0009] Accordingly, there is a need for systems and methods for amplitude and phase matching of the received signal with the reconstructed signal in a communication system using layered modulation. The present invention meets these needs.

SUMMARY OF THE INVENTION

[0010] Layered modulation reconstructs the upper layer signal and removes it from the received signal to leave a lower layer signal. Lower layer signal demodulation performance requires good signal cancellation, which in turn requires the reconstructed signal to include accurate amplitude and phase effects from signal

propagation path, filters and low noise block (LNB). Values of these parameters may change from receiver to receiver and therefore must be estimated at each receiver.

[0011] Embodiments of the invention utilize a technique to estimate the multiplicative relationship of magnitude and phase components between received and synthesized upper layer signals. These attributes will be multiplied to the signal synthesized from the satellite response, known transmitter and receiver filter characteristics, and estimated narrowband phase noise without additive white Gaussian noise (AWGN). The result of this multiplication is a high-fidelity representation of the upper layer signal which greatly enhances the cancellation performance. In addition, the required computational processing to implement the invention is minimal.

[0012] A typical method of the invention includes receiving a layered modulation signal including an upper layer signal and a lower layer signal in noise and interference, demodulating and decoding the upper layer signal from the received signal, estimating an upper layer amplitude factor and an upper layer phase factor from the received layered modulation signal. A substantially ideal upper layer signal is reconstructed from the demodulated and decoded upper layer signal including matching an ideal amplitude and an ideal phase by respectively applying the upper layer amplitude factor and the upper layer phase factor to the reconstructed ideal upper layer signal. Finally, the reconstructed ideal upper layer signal is subtracted from the received signal to produce the lower layer signal for processing.

[0013] A typical apparatus of the invention includes a signal processor for demodulating and decoding an upper layer signal from a received layered modulation signal wherein the received signal includes the upper layer signal and a lower layer signal in noise and interference. An estimator provides an estimate of an upper layer amplitude factor and an upper layer phase factor from the received layered modulation signal. A synthesizer reconstructs a substantially ideal upper layer signal

from the demodulated and decoded upper layer signal including matching an ideal amplitude and an ideal phase by respectively applying the upper layer amplitude factor and the upper layer phase factor to the reconstructed ideal upper layer signal. Finally, the lower layer signal is produced for processing by subtracting the reconstructed ideal upper layer signal from the received layered modulation signal with a subtractor.

[0014] Typically, the received layered modulation signal is a multiple phase shift keyed (PSK) signal in each layer and can comprise separate non-coherent modulated signal layers. Embodiments of the invention can estimate the upper layer phase factor from a mean vector of a distribution of one or more constellation nodes of the upper layer signal from the received layered modulation signal. The upper layer phase and amplitude factors can be estimated from a plurality of constellation nodes of the upper layer signal.

[0015] Furthermore, a transmission characteristic map can also be applied to improve the estimates of the upper layer amplitude and phase factors. The transmission characteristic map can comprise AM-AM maps and AM-PM maps characterizing effects of the transmission path. For example, the transmission characteristic map can represent a non-linear distortion map of amplifier characteristics of the transmission path, such as the effect of a travelling wave tube amplifier (TWTA) in a satellite transmission path.

BRIEF DESCRIPTION OF THE DRAWINGS

[0016] Referring now to the drawings in which like reference numbers represent corresponding parts throughout:

[0017] FIGS. 1A-1C illustrate a layered modulation signal constellation for an exemplary QPSK signal format;

[0018] FIGS. 2A and 2B illustrate a signal constellation of a second transmission layer over the first transmission layer before and after first layer demodulation;

[0019] FIG. 3 is a block diagram for a typical system implementation of the present invention;

[0020] FIGS. 4A and 4B illustrate the problem and the solution, respectively, using QPSK as an example;

[0021] FIG. 5 is an overview of the layered modulation reception process including the legacy receiver processes;

[0022] FIG 6 is a flowchart of the signal cancellation process; and

[0023] FIGS. 7A and 7B illustrate a general solution for amplitude and phase matching between received and reconstructed signals.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

[0024] In the following description of the preferred embodiment, reference is made to the accompanying drawings which form a part hereof, and in which is shown by way of illustration a specific embodiment in which the invention may be practiced. It is to be understood that other embodiments may be utilized and structural changes may be made without departing from the scope of the present invention.

1. Overview

[0025] FIGS. 1A - 1C illustrate the basic relationship of signal layers in an exemplary layered modulation transmission. FIG. 1A illustrates a first layer signal constellation 100 of a transmission signal showing the signal points or symbols 102. FIG. 1B illustrates the second layer signal constellation of symbols 104 over the first layer signal constellation 100 when the layers are coherent. FIG. 1C illustrates a

second signal layer 106 of a second transmission layer over the first layer constellation where the layers are non-coherent. The second layer 106 rotates about the first layer constellation 102 due to the relative modulating frequencies of the two layers in a non-coherent transmission. Both the first and second layers rotate about the origin due to the first layer modulation frequency as described by path 108.

[0026] FIGS. 2A - 2B illustrate a signal constellation of a second transmission layer over the first transmission layer. FIG. 2A shows the constellation 200 before the first carrier recovery loop (CRL) and FIG. 2B shows the constellation 200 after CRL. In this case, the signal points of the second layer are actually rings 202. Relative modulating frequencies cause the second layer constellation to rotate around the nodes of the first layer constellation. After the second layer CRL this rotation is eliminated. The radius of the second layer constellation is determined by its power level. The thickness of the rings 202 is determined by the carrier to noise ratio (CNR) of the second layer. As the two layers are non-coherent, the second layer may be used to transmit analog or digital signals.

[0027] FIG. 3 is a block diagram for a typical system 300 implementation of the present invention. Separate transmitters 316A, 316B, as may be located on any suitable platforms, such as satellites 306A, 306B, are used to non-coherently transmit different layers of a signal of the present invention. Uplink signals are typically transmitted to each satellite 306A, 306B from one or more transmit stations 304 via an antenna 302. The layered signals 308A, 308B (downlink signals) are received at receiver antennas 312, 320, such as satellite dishes, each with a low noise block (LNB) 310, 318 where they are then coupled to integrated receiver/decoders (IRDs) 314, 322. Because the signal layers may be transmitted non-coherently, separate transmission layers may be added at any time using different satellites 306A, 306B or other suitable platforms, such as ground based or high altitude platforms. Thus, any composite signal, including new additional signal layers will be backwards

compatible with legacy receivers which will disregard the new signal layers. Of course, non-backwards compatible applications are also possible in which both IRDs 314 and 322 are layered modulation IRDs, capable of receiving more than one signal layer. To ensure that the signals do not interfere, the combined signal and noise level for the lower layer must be at or below an allowed threshold level for the upper layer.

[0028] To receive layered modulation signals the upper layer signals must be reconstructed to cancel them from the total signal for lower layer signal processing to occur. Further, the performance of lower layer demodulation depends on the signal cancellation accuracy. The reconstructed signal should optimally match the received signal in overall amplitude and phase. Therefore, amplitude and phase errors in the reconstructed signal at the point of signal cancellation need to be estimated. The core of this invention includes techniques to optimally estimate a relative amplitude and phase between the received and reconstructed signals.

2. Amplitude and Phase Matching

[0029] FIG. 4A illustrates the problem which requires amplitude and phase matching, using QPSK as an example. FIG. 4A illustrates the QPSK constellation 400 before constellation amplitude and node phase compensation. All of the four triangles 402 (defining the phase error, θ_e) are identical. Embodiments of the invention can be applied to other modulation formats as well, such as 8PSK and 16QAM. The four nodes 404, represented by circles in FIG. 4A are ideal symbol locations after upper layer demodulation. They have a reference magnitude of one and respective phase angles of $\pi/4$, $3\pi/4$, $5\pi/4$, and $7\pi/4$. However, in real applications before they are converted by the analog to digital (A/D) converter at the receiver, the ideal symbol nodes 404 will have shifted in magnitude and phase with uncompensated and uncalibrated factors as represented by the actual nodes 406. FIG. 4B illustrates the collapsed ideal nodes 404 and the actual nodes 406.

[0030] The uncalibrated power represents an unknown magnitude scaling factor to the signal into the receiver (e.g., the set-top box). The low noise block (LNB), filters and other factors prior to the receiver typically introduce a phase distortion factor. These distortions should be included in the reconstructed upper layer signal to improve signal cancellation performance. As described above, FIG. 4A models these unknown distortions. The magnitudes of the received nodes for the upper layer signal are different from the assumed value of one, and are modeled by a relatively constant but unknown scaling factor, a_e. The received nodes for the upper layer signal are also offset from the ideal nodes by an equal but unknown phase adjustment amount, the phase error, θ_e. The signal is also corrupted with noise, interference and a second signal, represented by concentric circles around the nodes in FIGS. 4A and 4B. However, knowledge of each symbol node of the QPSK upper layer signal is available from forward error correction (FEC) decoding.

3. Exemplary Receiver Embodiment

[0031] FIG. 5 is a block diagram of the layered modulation reception apparatus. As shown, a receiver or integrated receiver/decoder (IRD) 500 embodiment of the invention estimates the upper layer power and phase which is used to re-scale the re-modulated signal, before the signal is subtracted from the received signal to leave only the lower layer signal. The signal 502 is received and the upper layer is demodulated by the demodulator 504. The demodulated signal 506 is then decoded (e.g., forward error correction decoding) by the decoder 508 to produce symbols 510 which are then communicated to the upper layer transport 512 for further processing and presentation to a viewer. The demodulator 504 and decoder 508 can be referred to in combination as a signal processor for processing the received signal. The foregoing processes encompass the functions of a legacy receiver decoding only the upper layer of the incoming signal 502 in cases of backwards compatible applications.

[0032] The lower layer of the incoming signal 502 requires further processing to decode. An ideal upper layer signal is generated by a synthesizer or remodulator 514. The remodulator 514 receives the upper layer timing and carrier 516 from the upper layer demodulator 504 and the upper layer symbols 510 output from the decoder 508. To enhance production of the ideal upper layer signal, the remodulator 514 can also receive input from a pulse shaping filter 518 and a non-linear distortion map 520 (which models transmission characteristics applied to the signal by elements such as the travelling wave tube amplifiers [TWTA] of the satellite).

[0033] A key element of the present invention comprises an estimator 522 which receives the incoming signal 502 and estimates an upper layer amplitude and phase factor. The factor is supplied to the remodulator 514 to further improve accurate reproduction of the ideal upper layer signal and benefit recovery of the lower layer.

[0034] The ideal upper layer signal is communicated to a subtractor 524 where it is subtracted from the incoming signal 502 which has been appropriately delayed by a delay function 526 to account for the processing time of the upper layer demodulator 504 and the remodulator 514. The output of the subtractor 524 is the lower layer signal which is communicated to the lower layer demodulator 528 and decoder 530 to produce the lower layer symbol output 532 which is ready to be processed by the lower layer transport for presentation.

4. Amplitude and Phase Matching for Constant-Envelope Signals

[0035] FIG. 6 is a flowchart of the signal cancellation process 600. As the received signal enters into the IRD 502 at block 602, the upper layer signal is first demodulated and decoded as described above at block 604. Meanwhile, an ideal upper layer signal is synthesized at block 606 with the decoded symbols 608 and other waveform parameters 610 derived from block 602. The synthesized signal is then mapped with TWTA AM-AM and AM-PM curves at block 612, which are positioned with a

suitable operating point estimate 614 obtained from the local upper layer demodulator 604 or downloaded from broadcast center, shown at block 616.

[0036] Typically, the TWTA performance maps will comprise measurements of the output amplitude modulation versus the input amplitude modulation (the AM-AM map) and the output phase modulation versus the input amplitude modulation (the AM-PM map). In the present invention, the received signal represents the amplifier output (plus lower layer signal, interference and noise) and the generated ideal signal represents the amplifier input. The maps are used to determine the effect of the TWTA on the signal and simulate those effects in the layer subtraction to yield a more precise lower layer signal. These performance maps are used to facilitate and/or improve reception of different layers of a system using a layered modulation transmission scheme.

[0037] Estimation of the operating point and AM-AM and AM-PM mapping are further discussed in U.S. Patent Application 10/165,710 filed June 7, 2002, by Ernest C. Chen and entitled "SATELLITE TWTA ON-LINE NON-LINEARITY MEASUREMENT", and Utility Application Serial No. 09/844,401, filed April 27, 2001, by Ernest C. Chen, entitled "LAYERED MODULATION FOR DIGITAL SIGNALS," which are both incorporated by reference herein.

[0038] The TWTA-mapped signal and the received signal are used to estimate the overall amplitude and phase factors in block 618. The mapped TWTA signal is then matched to the received signal in amplitude and phase at block 620. Finally, the corrected signal is subtracted at block 622 from the received signal, which has been properly delayed for timing alignment at block 624, to reveal the lower layer signal at block 626.

[0039] The key process of the present invention lies in referencing the received signal to the reconstructed signal, as in blocks 618 and 620. Referring back to FIG.

4A, a ratio is formed between the received signal and its decoded node signal. FIG. 4B shows the distribution of these complex ratios in effective additive noise; division by the decoded node signal collapses the received signals from all QPSK nodes to a single node near the horizontal axis. The mean of this distribution is the center of the concentric circles that represent the noise distribution. The mean vector is the estimate for signal matching purposes. The estimate vector consists of an amplitude, a_e, and phase, θ_e. The mathematical derivation is shown as follows.

r_i is the received signal for the i-th upper layer symbol in an effective noise;

n_i is the effective noise associated with r_i;

$\theta_{(i)}$ is the decoded phase for the i-th symbol;

N_s is the number of signal symbols processed;

a_e is the amplitude scale error to be estimated;

θ_e is the angular error to be estimated; and

$\theta_{(i)} \in \left\{\frac{\pi}{4}, \frac{3\pi}{4}, \frac{5\pi}{4}, \frac{7\pi}{4}\right\}$ for QPSK. Other modulation forms may be processed with a generalized solution discussed in the next section.

The received signal after carrier recovery can be modeled as:

$$r_i = a_e \exp\left(j\left(\theta_{(i)} + \theta_e\right)\right) + n_i, \text{ where } i = 1,\ldots,N_s. \quad (1)$$

Removing the decoded symbol phase yields:

$$r'_i \equiv \frac{r_i}{\exp\left(j\theta_{(i)}\right)} = a_e \exp\left(j\theta_e\right) + n'_i; \quad (2)$$

where $n'_i = \frac{n_i}{\exp\left(j\theta_{(i)}\right)}$. n_i and n'_i have zero mean and the same variance. The estimated complex amplitude and phase scale factor is formed by averaging over r'_i as follows.

$$r_0 \equiv avg\left\{r'_i\right\} = \frac{\sum_{i=1}^{N_s} r'_i}{N_s} \equiv \hat{a}\exp\left(j\hat{\theta}_e\right). \tag{3}$$

The amplitude and phase error estimates are:

$$\hat{a}_e = abs\left(r_0\right); \text{ and} \tag{4}$$

$$\hat{\theta}_e = angle\left(r_0\right). \tag{5}$$

[0040] The preceding analysis shows that the estimated residual phase $\hat{\theta}_e$ will be zero if the signal phase has been precisely followed with carrier recovery loop, etc. $\hat{\theta}_e$ "sweeps up" residual phase errors due to carrier recovery inaccuracy and other errors.

[0041] As shown by amplitude and phase matching 620 operation in FIG. 6, the estimated amplitude and phase factors form a complex multiplier to the reconstructed signal for subtraction from the delayered received signal to optimally reveal the lower layer signal.

5. Amplitude and Phase Matching for General Signals

[0042] FIGS. 7A and 7B illustrate a general solution for amplitude and phase matching between received and reconstructed signals that are not restricted to QPSK. As shown in FIG. 7A, all triangles 702 are similar with ratios $a_1, a_2, \ldots a_k$, etc. Thus, the technique, described above with respect to QPSK, can be easily extended for use

with a reconstructed signal which varies in amplitude due to unequal signal node amplitudes, variations due to inter-symbol interference prior to matched filtering, satellite non-linear response, etc. The reconstructed signals are shown collapsed in FIG. 7B after node phase compensation with unequal amplitudes between ideal nodes 704A and 704B as well as their respective received nodes 706A and 706B.

[0043] A general analysis for amplitude and phase matching of the present invention follows. This analysis degenerates to the preceding analysis when applied to a QPSK constellation which has identical magnitudes (amplitudes). For a general layered communication signal, the constellation symbols may utilize different amplitudes and phases. Thus,

$a_{(i)}$ is the amplitude of the i-th symbol over time;

$\theta_{(i)}$ is the phase of the i-th symbol over time; and

$s_{(i)} = a_{(i)} \exp\left(j\theta_{(i)}\right)$.

The received signal after the carrier recovery loop can be modeled as:

$$r_i = a_e a_{(i)} \exp\left(j\left(\theta_{(i)} + \theta_e\right)\right) + n_i \text{, where } i = 1, \ldots, N_s. \quad (6)$$

Removing the remodulated and re-encoded signal phase and weighting by the signal magnitude, similar to matched-filtering forms:

$$r'_i = \frac{a_{(i)} r_i}{\exp\left(j\theta_{(i)}\right)} = a_e a_{(i)}^2 \exp\left(j\theta_e\right) + n'_i \quad (7)$$

where $n'_i = \dfrac{a_{(i)} n_i}{\exp\left(j\theta_{(i)}\right)}$

and n_l and n'_l have zero mean. The estimated complex amplitude and phase scale factor is formed by summing over r'_l, normalized by the sum of the ideal powers as follows.

$$r_0 \equiv \frac{\sum_{l=1}^{N_s} r'_l}{\sum_{l=1}^{N_s} a_{(l)}^2} \equiv \hat{a}_e \exp\left(j\hat{\theta}_e\right). \tag{8}$$

As before, the amplitude and phase error estimates are:

$$\hat{a}_e = abs\left(r_0\right); \text{ and} \tag{9}$$

$$\hat{\theta}_e = angle\left(r_0\right). \tag{10}$$

Note that equation (8) reduces to equation (3) when all $a_{(l)}$ are equal. However, the general solution of equation (8) may be preferred even for *n*PSK signals since all received signal symbols have non-constant amplitudes prior to receiver matched filtering.

6. Alternative General Analysis for Amplitude and Phase Matching

[0044] An alternative approach to the problem, which results in the same solution as the preceding general solution can be found through vector analysis. The approach begins with the same mathematical model, but uses complex numbers to represent phases and magnitudes of the received symbols.

[0045] The problem is characterized in terms of a minimization process. Suppose $\underline{R}$ is the received signal vector and $\underline{X}$ is the reconstructed signal vector; the vectors consist of the associated time samples as their components. Both are column vectors

with length N_s, where N_s is the number of data symbols to be processed. A complex scalar factor z is to be estimated for multiplication to $\underline{X}$ later. The estimate is chosen to minimize the difference between $\underline{R}$ and $z\underline{X}$, or specifically, the norm-squared error: $(\underline{R} - z\underline{X})^H (\underline{R} - z\underline{X})$, where $(\,)^H$ is the Hermitian operator. The result is a least-square-error (LSE) solution:

$$z_{LS} = \frac{\underline{X}^H \underline{R}}{\underline{X}^H \underline{X}} \tag{11}$$

$\underline{X}^H \underline{X}$ is a scalar equal to the power of the reconstructed signal. z_{LS} is the complex correlation between received signal vector $\underline{X}$ and reconstructed signal vector $\underline{R}$, normalized by $\underline{X}^H \underline{X}$. Thus, z_{LS} is the complex correlation of the received signal vector and the reconstructed signal vector and normalized by a power of the reconstructed signal vector, identical to the previous solution expressed by equation (8).

[0046] This concludes the description including the preferred embodiments of the present invention. The foregoing description of the preferred embodiment of the invention has been presented for the purposes of illustration and description. It is not intended to be exhaustive or to limit the invention to the precise form disclosed. Many modifications and variations are possible in light of the above teaching.

[0047] It is intended that the scope of the invention be limited not by this detailed description, but rather by the claims appended hereto. The above specification, examples and data provide a complete description of the manufacture and use of the apparatus and method of the invention. Since many embodiments of the invention can be made without departing from the scope of the invention, the invention resides in the claims hereinafter appended.